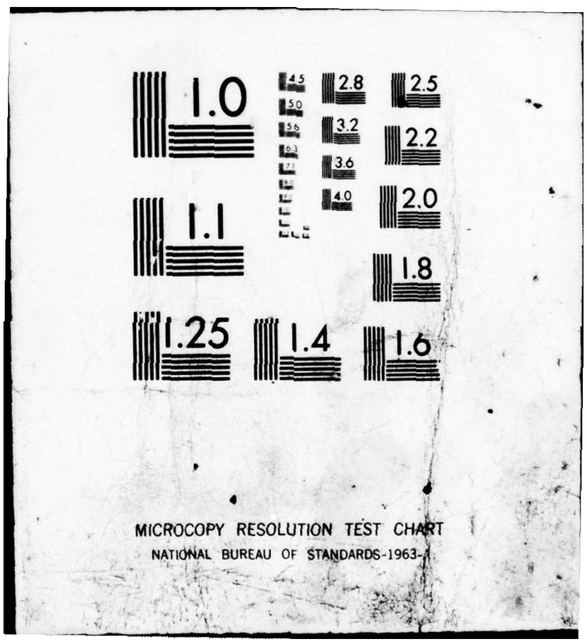
AD-A068 710 CARNEGIE-MELLON UNIV PITTSBURGH PA MANAGEMENT SCIENC--ETC F/G 5/1
A DATA ENVELOPMENT ANALYSIS APPROACH TO EVALUATION OF THE PROGR--ETC(U)
NOV 78 A CHARNES, W W COOPER, E RHODES N00014-76-C-0932

UNCLASSIFIED MSRR-432 NL

1 OF 1
AD
A068710

END
DATE
FILMED
7--79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

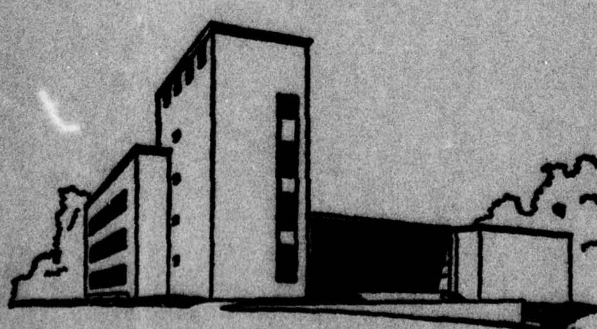
LEVEL III

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION

WILLIAM LARIMER MELLON, FOUNDER



79 05 14 104

14 MSRR-432

12 64p.

9 Management Sciences Research Report No. 432

6

A DATA ENVELOPMENT ANALYSIS APPROACH
TO EVALUATION OF THE PROGRAM FOLLOW THROUGH
EXPERIMENT IN U.S. PUBLIC SCHOOL EDUCATION.

by

10 A. Charnes*
W. W. Cooper**
E. Rhodes***

11 Nov[redacted]78

15 N00014-76-C-0932
√NSF-SOC76-15876

* The University of Texas
** Graduate School of Business Administration, Harvard University
*** School of Management, State University of New York at Buffalo

This research was partly supported by NSF Grant No. SOC76-15876 "Collaborative Research on the Analytical Capabilities of a Goals Accounting System." It is also supported by Project NR 1947-021, ONR Contract N00014-75-C-0616 with the Center for Cybernetic Studies, The University of Texas, and ONR Contract N00014-76-C-0932 At Carnegie-Mellon University School of Urban and Public Affairs. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

Management Sciences Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

403 426 Gm

ABSTRACT

alpha

A method called Data Envelopment Analysis (DEA) is used to decompose the efficiency of Decision Making Units (DMU's) into two parts: (1) a component resulting from managerial decisions and (2) a component resulting from constraints (called programs) under which management operates. The DEA approach accomplishes this by enveloping the input-output observations with extremal relations developed in terms of a specified nonlinear programming model (and/or its linear programming equivalent). Differences between the observations and the program specific envelopes -- called α-envelopes--are imputed to managerial inefficiencies. An inter-program envelope is then constructed from 2 or more such α-envelopes and used to identify "program" inefficiencies, which are the inefficiencies that remain after the previously determined managerial inefficiencies have been eliminated. Numerical illustrations accompanied by suggested tests of a probabilistic/information theoretic character are provided by means of recently released data from "Program Follow Through." Designed as a study of possible ways of reenforcing or extending Program Head Start - an ongoing pre-school program for disadvantaged children -- the Program Follow Through experiment provides data on agreed upon inputs and outputs for both PFT (Program Follow Through) and matched NFT (Not Follow Through) participants in various parts of the U.S. Only a subset of the variables from the Follow Through experiment are used. Hence the numerical example utilized here is best regarded as only illustrative. Although the results are adverse to PFT, the DEA approach also opens new ways of profiting from the results of such experiments by examining combinations of the underlying components. These kinds of possibilities are also described in this paper.

KEY WORDS

Efficiency

Program Efficiency

Managerial Efficiency

Decision Making Units

Program Follow Through

Educational Outputs

Mathematical Programming

Linear Programming

Duality Relations

Extremal Relations

Efficiency Frontiers

Isoquants

Production Possibility Surfaces

Information Measures

Regression

Simultaneous Estimation

1. BACKGROUND

In [10] and [23] we were concerned with developing ways to measure the efficiency--more precisely, the relative efficiency--of decision making units (DMU's) with special reference to public sector applications. These measures were, in general, to be secured from observations on input and output values that resulted from past decisions. The objective was to devise new methods for dealing with multiple outputs as well as multiple inputs such as are of interest for most public sector activities. We wanted our efficiency measures to be obtained directly from such input and output data (all of it) in an objective manner--i.e., without recourse to a priori weighting choices and like artifacts, such as price imputations from private markets to public sector activities, which ~~such as~~ have customarily been employed for evaluating public sector activities.

The developments in [10] and [12] naturally entailed a variety of new methods (e.g., for estimating extremal relations from empirical data) as well as new ways of unifying and using older concepts and methods (e.g., the different lines of research embodied in the works of Farrell [17] and Shephard [26]). These will not be examined here. We want to concentrate instead on extending the ideas of efficiency measurements by developing methods that are directed toward identifying the efficiency of DMU's under differing program possibilities. That is, we seek to identify the efficiencies that are possible under different public programs (e.g., different programs of public school education) and to distinguish these aspects of efficiency from the way the DMU's operating under these programs avail themselves of these opportunities.

How we propose to effect these separations from given observations will be developed in subsequent sections of this paper and illustrated in detail by

reference to data from an important experiment in U.S. public school education that has come to be known as "Program Follow Through." Here we may observe that such an evaluation depends in part on the ability to distinguish between "program efficiency and "management efficiency" where the latter refers to the efficiency of the DMU's operating under a given program. For if we fail to make such a distinction, we are in danger of faulting what may be a "good program" as a result of management (as distinct from program) inefficiency and, conversely, a "bad program" may be approved because of the efficiency of the DMU's operating under its restrictions.

The data of Table 1, as drawn from [10], will help us clarify what is intended in our characterization of DMU (managerial) efficiency. Here we are supposing that each of three DMU's produces a single unit of the same output by means of two inputs in the amounts x_1 and x_2, that are shown row by row, under the columns for DMU_1, DMU_2 and DMU_3. (We may also think of these x_1, x_2 values as the requirement per unit output as obtained from empirical data by norming each DMU's inputs by its total output.)

Evidently DMU_2 is not as efficient as DMU_1 and hence cannot be characterized as being efficient. For reducing $x_1 = 3$ to $\hat{x}_1 = 2$ while holding $x_2 = 2$ in order to produce DMU_2's one unit of output would bring this management into coincidence with DMU_1.

TABLE 1
AN ILLUSTRATION OF
DMU (=MANAGERIAL) EFFICIENCY

INPUT \ DMU NO.	1	2	3
x_1	2	3	4
x_2	2	2	1

We are assuming, of course, that all inputs and outputs have some "value."[1] That is, we assume that released resources can be used elsewhere and/or that an expanded output from the given inputs has some value. In other words we assume that the indicated reduction of x_1 represents a gain. However, this reduction neither uses all of the information of Table 1 (since DMU_3 is ignored) nor gives the measure of DMU efficiency we are seeking.

To obtain the desired measure of DMU efficiency we import the following formulation from [10] and [23]:

$$\max h_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

subject to

$$(1) \qquad 1 \geq \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \quad ; \; j=1, \ldots, n$$

$$u_r, v_i > 0; \; r=1, \ldots, s; \; i=1, \ldots, m.$$

The y_{ij}, x_{ij} values, which are all positive constants, represent observed amounts of $r=1, \ldots, s$ outputs and $i=1, \ldots, m$ inputs for each of the $j=1, \ldots, n$ decision making units (DMU's) that constitute a reference set. For each phase of this efficiency evaluation, one member of this set is singled out and represented in the functional as well as in the constraints. The resulting optimization yields a set of objectively determined weights[2] u_r^*, v_i^* which generate an optimal $0 \leq h_o^* \leq 1$ with $h_o^* = 1$ if and only if the thus distinguished DMU is efficient.

---

[1] But without specifying their numerical magnitudes in advance.

[2] The reasons for restricting these weights to positive values are set forth in the "Corrections" to [10].

We now apply what was previously said to determine such an h_o^* for DMU_2 in Table 1. This means that the data for DMU_2 will appear in the functional as well as the constraints so that for this single output case (with $y_j = 1$, for $j = 1, 2, 3$) we have:

$$\max h_o = \frac{1u}{3v_1 + 2v_2}$$

subject to

$$1 \geq \frac{1u}{2v_1 + 2v_2}$$

$$1 \geq \frac{1u}{3v_1 + 2v_2}$$

$$1 \geq \frac{1u}{4v_1 + 1v_2},$$

$$u, v_1, v_2 > 0.$$

As shown in [10] and [23] the formulation in (1) can be replaced by an ordinary linear programming equivalent for which calculations in this case. yield

$$u^* = 1, \; v_1^* = 1/6, \; v_2^* = 1/3.$$

This solution evidently satisfies all constraints since

$$\frac{u^*}{2v_1^* + 2v_2^*} = 1$$

$$\frac{u^*}{3v_1^* + 2v_2^*} = 6/7$$

$$\frac{u^*}{4v_1^* + 1v_2^*} = 1,$$

and it provides the optimal functional value with $h_o^* = 6/7$, the same as in the constraint for j = 2.

We may observe that this same calculation which established the inefficiency of DMU_2, also established that DMU_1 and DMU_2 are efficient. They, in fact, serve as the efficient references for establishing this value of $h_o^* = 6/7$ for DMU_2. This means that a suitable convex combination of the data for DMU_1 and DMU_3 provides an efficient reference point for establishing the efficiency of DMU_1. In fact this convex combination is formed from the data vectors to give,

$$\frac{5}{7}\begin{pmatrix}2\\2\end{pmatrix} + \frac{2}{7}\begin{pmatrix}4\\1\end{pmatrix} = \frac{6}{7}\begin{pmatrix}3\\2\end{pmatrix}.$$

The supposition is that efficiency requires DMU_2 to be on the line (the unit isoquant) generated from all convex combinations of the data for DMU_1 and DMU_3. To achieve a position on this unit isoquant, however, DMU_2 would have had to employ 6/7 of the amount of each of the two inputs utilized. See the contraction factor on the right hand side in the above expression. Conversely, we could multiply through by ~~6/7~~ 7/6 to obtain

$$\frac{5}{6}\begin{pmatrix}2\\2\end{pmatrix} + \frac{2}{6}\begin{pmatrix}4\\1\end{pmatrix} = \begin{pmatrix}3\\2\end{pmatrix}$$

which means that with these input amounts, an efficient DMU_2 would have augmented its output from $y_2 = 1$ to $\frac{5}{6} + \frac{2}{6} = \frac{7}{6}$ units. In other words, h_o^* measures the amount of resource conservation that is required for efficient production of a given output and its reciprocal measures the amount of output augmentation that is required for efficient utilization of given inputs. The measures in any case are by reference to subsets of DMU's which utilize the same inputs and outputs in a relatively efficient manner. Witness, e.g., DMU_1 and DMU_3 which are both efficient in the above illustration.

These ideas go over to multiple outputs and inputs and, as we shall see in our Program Follow Through example, they can be extended to inter-program efficiency comparisons as well. Before effecting our extensions in these directions, however, we can usefully conclude the present section by emphasizing that our efficiency measure differs in important ways from other commonly employed measures such as indexes of productivity, etc. The latter, for instance, generally proceed by reference to only one input and one output at a time without attempting to distinguish efficient from inefficient operations whereas our efficiency measure considers all inputs and outputs that are represented as in (1) and then measures possible gains from altering the input and output combinations utilized. In the latter connection our efficiency measure is intended to have the operational significance we have just exhibited for input reduction or output augmentation whereas, in general, such significance cannot be assigned to the usual indexes of productivity, etc.[1]

---

[1] E.g., indexes of Laspeyre or Paasche variety.

2. ESTIMATION OF PRODUCTION AND EFFICIENCY RELATIONS

The above efficiency comparisons assume certain underlying relations which may be summarized by means of concepts from the field of production economics. In particular we draw on the concept of a production function as developed for the case of a single output. Such a function is defined as an extremal relation, by which we mean that the output value is assumed to be maximal for any inputs that may be specified.

In our case this production function is piecewise linear and with returns to scale (and marginal rates of substitution) that are consequently piecewise constant.[1] Note, however, that these pieces are ~~also~~ determined from the data. They are also accompanied by an adjustment procedure that ensures attainment of the efficient surface even in the case of multiple outputs. See [10] and [23].

This situation differs from the usual approach to the study of production functions (and related efficiency surfaces) in empirical economics. The latter, we may say, generally proceeds by methods such as least-squares regressions or simultaneous estimation systems which emphasize averages or like measures of "central tendency." These approaches therefore fail to match the methods of estimation with the underlying theoretical constructs which, as we have just noted, proceed by means of extremal relations. This means that a variety of possibilities for eliminating waste and inefficiency are concealed from view[2] by virtue of the statistical methods employed.

[1] This has since been generalized to functions that are piecewise Cobb-Douglas, piecewise translog, etc., and which may exhibit both increasing and decreasing returns to scale in their different pieces. See [12]. Utilization here of notions such as increasing and decreasing returns to scale, etc., would, however, seem to be pushing beyond the boundaries of what the test data we are using from Program Follow Through will stand.

[2] This deficiency is present in almost all of the econometric studies which have been directed to issues of energy policy. (For a discussion of other shortcomings of approaches used in such studies see A. Charnes, W.W. Cooper and A. Schinnar "Transforms and Approximations in Cost and Production Function Relations." University of Texas at Austin, Center for Cybernetic Studies, Dec., 1976.)

We have already provided one way of locating such inefficiencies and illustrated its application by reference to Table 1, and we shall shortly introduce additional extensions of these ideas. Before doing so, however, we need to note that our approach involves interpretations and approaches to data treatment that differ from the ones that have usually been employed in statistical-econometric investigations of areas like educational policy, etc. The latter may be referred to as utilizing a "prediction approach." This is intended to suggest that in such approaches one applies statistical regressions[1] or simultaneous estimation techniques to all of the data. The resulting relations are then used to predict further behavior on the assumption that decision makers will continue at past levels of efficiency.

In part this approach is contingent on the methodologies (such as least-squares) which are employed.[2] In part it is dependent on the kinds of data treatments utilized. In general no data adjustments are effected for purposes of distinguishing efficient from inefficient operations and hence no prediction is possible beyond the supposition that future behavior will continue to generate observations with similar mixes of efficiency and inefficiency.

Our approach provides ways of distinguishing the relative efficiency in the observations that have been generated from past DMU behavior. Accordingly we can also adjust the data to provide the extremal relations that are needed

---

[1] Extensive use of such regression and regression related techniques was made in the study of Program Follow Through that was conducted by Abt Associates. See [1] and [2].

[2] A discussion may be found in [23] which ranges from critical evaluation of the one-dependent-variable-at-a-time assumption of standard regression approaches and includes the limited ability of presently available simultaneous estimation models and related statistical methods for dealing with the large numbers of variables and relations that are present in the kinds of applications we are considering.

to characterize the production functions and related efficiency surfaces which are prescribed by our underlying theory.[1]

Our approach may be distinguished from others by referring to it as a "control prediction" approach. By this we mean that the data adjustment and estimation techniques will not yield good predictions unless suitable controls are applied. Thus in the Follow Through Program study, for example, we may find that observed input and output values contain elements of inefficiency. Although validated--e.g., by reference to the behavior of relatively efficient subsets of DMU's--there is nevertheless an assumption that these inefficiencies may be identified and eliminated.

Evidently something more than simply a prediction via relations derived from past observations is involved. We need to emphasize this since our approach also admits of simultaneous estimation of (multiple) output and input relations that permit still further improvement by means of substitutions between the various amounts of inputs and outputs utilized. Unlike the prediction approaches that we previously described[2] ~~, however,~~ the results of such substitutions have been previously adjusted for possible departure from the efficiency frontiers.

-- pure [illegible] description

---

[1] A discussion of ways in which these production functions (and efficiency surfaces) differ from others that have been used in economic analysis may be found in [10] and [23].

[2] Summers and Wolfe in [30], for example, utilize statistical regression techniques to reassign staff to different duties in order to achieve improvements in efficiency. In addition to the critique supplied in [5], we also need to underscore that these regressions continue to retain whatever mixes of efficiency and inefficiency were present in the data from which they were derived, and hence they differ from the extremal relations estimates we shall be using.

Because we want to emphasize the efficiency measurement and data adjustment processes in our Program Follow Through illustration, we shall not explore these further substitution possibilities in the present paper.[1] In any case we need machinery beyond that utilized in the pure prediction approaches in order to ensure that the wanted efficiencies will be forthcoming. For present purposes we may think of these as being identified by means of a field examination such as would accompany a "comprehensive audit."[2] Then we may think of our approach as providing the basis of what is termed an "analytic review" in audit practice. In such an a priori review, past data (and other pertinent information) are utilized to identify places where an auditor should look for possible problems and improvements. Although the models and approaches we shall be suggesting will have value for such a review, a full verification of our "control predictions" will generally require field examinations or other additonal modes of identification of possible sources of inefficiency.

---

[1] The values for the coefficients in these extremal relations are available (without extra computation) from the duals to the linear programming equivalents to (1). See [10] and [23].

[2] Briefly such comprehensive audits extend the usual concept of financial audit (e.g., CPA attest audits) to other aspects of management. See [13] and [31].

3. PROGRAM EFFICIENCY AND DATA ENVELOPMENT ANALYSIS

We now want to extend the above ideas to enable us to distinguish program from managerial efficiency in the different reference sets of DMU's we shall be studying. We therefore introduce the following extension of (1):

$$\max h_o^\alpha = \frac{\sum_{r=1}^{s_\alpha} u_r^\alpha y_{ro}^\alpha}{\sum_{r=1}^{m_\alpha} v_i^\alpha x_{io}^\alpha}$$

(2) subject to

$$1 \geq \frac{\sum_{r=1}^{s_\alpha} u_r^\alpha y_{rj}^\alpha}{\sum_{i=1}^{m_\alpha} v_i^\alpha x_{ij}^\alpha} ; \quad j=1, \ldots, m_\alpha$$

$$u_r^\alpha, v_i^\alpha > 0; \; r=1, \ldots, s_\alpha \; ; \; i=1, \ldots, m_\alpha ,$$

where $\alpha=1, 2, \ldots, k$, respectively, indexes the sets which are of interest.

Within each such set we will, of course, have the same efficiency measurement situation as before -- viz., $0 \leq h_o^{*\alpha} \leq 1$ with $h_o^{*\alpha} = 1$ if and only if the DMU being evaluated relative to the α^{th} set of DMUs is efficient. Now, however, we want to extend these ideas so that we can apply them across the sets $\alpha = 1, 2, \ldots, k$ in order to examine the relative efficiency of the sets themselves.

For this comparison, we shall require common outputs and inputs for the reference sets. Then, provisionally, we may think of this as a comparison between each of $\alpha = 1, 2, \ldots, k$ "technologies" in order to determine their varying degrees of efficiency for converting common inputs into common outputs.

Each such technology provides a "boundary" to the set of production possibilities under the usual assumptions of economic theory. We shall be dealing with direct inferences from empirical data, however, and so we will not be able to assume that all DMUs attain these boundaries. Furthermore, unless a knowledge of these boundaries is objectively available from some a priori source, we shall only be able to establish relative rather than absolute efficiency ratings by reference to the most efficient members of the respective reference sets. That is, these efficient subsets of DMUs will be used to establish the relative efficiency boundaries which we shall refer to as "envelopes" in order to emphasize these (and other) departures from the usual assumptions of economic theory and methods of empirical inquiry.

Before effecting our across-envelope comparisons, we shall bring each DMU onto the envelope for its reference set in the manner set forth in [23]. We shall mainly be concerned with behavior such as the behavior of educational institutions in the public sector -- where perfectly competitive market forces are not ordinarily given free play. Nevertheless, in a rough

sort of analogy we may think of these adjustments as corresponding to that part of competitive theory in which each DMU is forced to become as efficient as the most efficient of its competitors as a condition for survival.[1] Note, however, that this is not an assumption in our case since we actually adjust the observations in this manner. Thus we are able to effect these across-envelope comparisons on the basis of data which are adjusted so that all DMUs are as efficient as the most efficient among them. The resulting comparisons across these envelopes will then be used to rate the respective efficiencies of these envelopes.

Naturally, we shall need to utilize analytic methods that enable us to effect comparisons between the different distributions of efficiencies both within and across each such set. This will be done by a variety of methods, including uses of the "divergence" measure of information theory for determining the "distance" between different distributions. The usual tests of significance may then be applied, but it is important to emphasize that we are proceeding in an order that is the reverse of the usual one. That is, unlike the situation in which one wants to test an underlying theory, we are here using that theory to bring the observations onto the envelope that serves as the efficiency frontier in each set. Only after this has been done are the tests of significance to be applied.

As we shall see, this approach considerably simplifies the kinds of statistical models and methods that may be employed and it opens a variety of applications for policy evaluations and controls that are not available

[1] We shall also refer to our envelopes as "efficiency frontiers" even though we do not make the usual profit maximizing (incentive) assumption that the most efficient DMUs always effect the best choice that technology makes possible. See section 5, below.

from more customary approaches.[1] In any event we need to distinguish this approach which we shall refer to as Data Envelopment Analysis (DEA) and which we now try to motivate in the following way: Suppose we have two different programs that might be used in public education. Each program has the same (multiple) output objectives and utilizes the same inputs as, for instance, in the experiment on Program Follow Through (PFT) that we shall shortly examine. In deciding whether PFT is better than its alternative, Non Follow Through (NFT), we need to allow for a variety of possibilities in view of the fact that the observations for each of PFT and NFT contain deviations that can reflect decisions which fall short of what each program admits.

By distinguishing between program and managerial (= decision making) efficiency, our DEA approach is directed toward evaluating a variety of policy possibilities that need to be considered. As already noted, it enables us to distinguish between managerial and program efficiency so that, inter alia, we can determine whether program comparisons entail different degrees of managerial efficiency in the data sets, or whether allowances should be made for different degrees of DMU efficiency before effecting program evaluations. See section 5, below. Furthermore, the DEA approach singles out the more efficient DMUs for possible sutdy en route to setting standards and other types of controls within any such program. It also opens the possibility of synthesizing entirely new programs by identifying subsets of across-program DMU's (for which the envelopes intersect) as a possible source for forming new program combinations that are better than any of the originally identified programs. Of course, still other possibilities become available and, in any case, the DEA approach helps us to distinguish good programs which might be badly managed from worse programs that appear to be better because of management rather than program capability. It is this latter aspect of DEA which we shall emphasize in what follows, but we shall also at least indicate some of these other possibilities along the way.

---

[1] For a discussion of the meager results obtained to date from

4. PROGRAM FOLLOW THROUGH BACKGROUND

We shall illustrate these DEA ideas by reference to a body of data[1] that have recently become available from a very important experiment in U.S. public primary school education known as Project Follow Through (PFT).[2] Before commencing with the specifics of our analysis of the data from the Project Follow Through experiment, however, we briefly consider its history and development. It was conceived in the late 1960's as a Federally sponsored program charged with providing remedial assistance to educationally disadvantaged early primary school students.

To a large extent PFT was developed in response to perceived needs for furthering the objectives and accomplishments of the well known Project Head Start.[3] In fact, a major justification for Follow Through was to supplement

unusual courtesy, especially for data on public school education, which we herewith gratefully acknowledge. We are also grateful to Mary Kennedy, the Program Follow Through Project Officer at the U.S. Office of Education, for sending us the U.S. Office of Education Reports that are also referenced in [2].

[2] The study itself is known as <u>Education as Experimentation: A Planned Variation Model</u>. See the discussions in the U.S. Office of Education reports listed in [2].

[3] Project Head Start was designed as an early childhood <u>pre-school</u> intervention program aimed at bringing about significant cognitive and non-cognitive gains among disadvantaged children. When subsequent studies indicated that Project Head Start effects were not sustained after its participants entered primary school, Project Follow Through was one suggested corrective measure--the idea being that special attention in the first few grades would lend reenforcement to what Project Head Start had previously initiated. See the discussions on pp. 158 - 159 in Vol. IIA of the U.S. Office of Education reports in [2].

public schools did not articulate sufficiently with the "Head Start goals, curricula, and objectives in the early grades so that these children would maintain or accelerate their pre-school achievement."[1] Follow Through was envisioned as an answer. As the same time Follow Through was also to be a "community action program," going well beyond the classroom in providing for community services such as nutrition programs, social, medical, and dental assistance, and even psychological counseling service.

The academic portion of Follow Through could be interpreted as a form of Head Start which moved the latter from pre-school into the elementary grades at the level of kindergarten through third grade. Initially conceived as a program involving some 200,000 children, Follow Through received enthusiastic endorsements from a variety of educational authorities. For instance R. L. Eghert[2] writes:

> (I)ts design stemmed from the conviction that sufficient improvements could be effected in the institution serving children that children's development would be so markedly superior as to be readily demonstrated on measures of achievement, cognition, self-concept, social maturation, and capacity to function independently. Follow Through's design was born also from the conviction that unless such substantial differences were manifest, the really massive increases in spending that would be required could not be justified. In view of the results reported by Miller, Engelman, Gordon, and others in the January, 1968 meetings of prospective sponsors, this conviction did not seem unrealistic, assuming that programs developed in small scale settings could be implemented on a larger scale in a number of communities.

---

1) See, e.g., Stanford Research Institute (SRI) [29], pp.2-3.

2) Quoted in [1] pp.A-7 and A-8.

Unfortunately for those anticipating a large scale primary school action program, a series of developments occurred between Follow Through inception and funding which resulted in a change of Federal policy and a reduction in annual funding to only $15 million dollars from an originally proposed $200 million dollars.[1] This reduction in funding caused a rethinking in which the proposed massive application was converted into an "experimenal study" with the latter to be executed in an approach referred to as "planned variation."

The idea was to utilize an experimental design approach or at least as much of an approximation to these canons of classical statistics as one is likely to be able to secure in a field like educational policy.[2] Nevertheless, within these limits of the "planned variation model", the enacted Follow Through Program was to be formulated around a collection of specifically identified approaches to treating the

---

[1] The funding for this study is reported to have been as follows:

Academic Year (19-)	Funding ($\$10^6$)
67 - 68	3.75
68 - 69	11.25
69 - 70	32.00
70 - 71	70.30
71 - 72	69.00
72 - 73	63.06
73 - 74	50.62
74 - 75	52.85
75 - 76	55.42
76 - 77	59.00
Total	467.25

Source: p. 21 in W. Haney, The Follow Through Planned Variation Experiment, Vol. 5; The Follow Through Evaluation: A Technical History, prepared for the Office of Planning, Budgeting and Evaluation of the U.S. Education Department of HEW by The Huron Institute, Cambridge, Massachusetts, August, 1977.

[2] This is, of course, only a recognition of the particular susceptibility of education, and especially education in the early grades, to emotions, pressures and other impediments to purely scientific studies.

compensatory education problems of disadvantaged children. These program variations were each associated with "sponsors" (e.g. sponsors headquartered at local universities or research institutes) who were to (1) provide the basic form and content of one particular "planned variation" and (2) work with designated local school districts in implementing the indicated variation. See Tables A-1 and A-2 in the Appendix.

Conformance with the above conditions was to be a requirement for Federal funding (and related resource advantages) and, further, this was extended to a directive that each school district supply a Non-Follow Through as well as a Follow Through candidate group. Naturally, allowance was made for periodic reports and analyses to facilitate study of these various programs, and competent statistical (and other) consultants were retained for effecting analyses of the resulting data. The results from these analyses were so mixed and subject to dispute and challenge, however, that we confine ourselves to a PFT versus NFT comparison without reference to the variations in the assorted Follow Through approaches identified with these different sponsors.[1]

---

[1] See [1] and [2] for a detailed treatment of the various Follow Through sponsor performances.

Another difficulty arises in that Follow Through provided a variety of social as well as educational services to the community. In fact, because of its original tie to the Community Action Program (CAP) arm of OEO (the U. S. Office of Economic Opportunity), the following four community services were mandated, in varying degrees, for all of the Follow Through Programs:

(1) Medical and dental services
(2) Nutritional programs
(3) Social service programs
(4) Guidance and psychological services.

Being community based and not attached to specific academic programs, it was not possible to determine the differential effects, if any, of these nonacademic activities of the Follow Through programs. Thus, we, like the other analysts, will simply ignore these parts of the Program in order to focus on only the academic portions of Follow Through.

Within the latter limits, certain attractive features emerge for our purposes. For one thing, the Follow Through study is almost unique among programs of its size in that all of the sites administered the same core battery of tests and measurements for the proposed national evaluation. This included the NFT as well as the PFT segments. Moreover, the former, i.e., the NFT sites, were selected to obtain matched comparison sets of supposedly comparable students. The PFT results could thereby be matched to comparable control populations rather than being confined only to comparisons with some supposedly general aggregate national norm. While this matching was not completely carried out in all detail, it at least provides a better basis than most of the

other quasi-experimental designs[1] of "planned variation" genres in educational policy.

5. SELECTION OF VARIABLES

In our opening section, we indicated some of the properties of our proposed DEA approach to program and managerial efficiency measurement. Now we might indicate others. Note, for instance, that the above matching presents certain difficulties that are not encountered in the classical (natural science) models of experimental design and which therefore require specific attention.

As a case in point we might consider the problem of managerial (= decision making) efficiency as it might be distributed between DMU's in PFT and NFT. Differences in decision making efficiency need to be allowed for since, evidently, a "good program" may be "badly managed," and _vice versa_, so that one needs some way of identifying this possible source of contamination in arriving at a "program" evaluation.

If these were profit making entities one might -- at least in principle -- use dollar scalarizations for both inputs and outputs in order to effect a matching for efficiency, possibly in the original experimental design. No such _a priori_ basis is available here, however, and so our DEA procedures are applied "after the fact," so to speak, to eliminate such managerial inefficiencies en route to effecting the wanted PFT-NFT comparisons.

Since we want to focus on the concepts and adjustment methodologies

---

[1] Cf., e.g., Campbell and Stanley [6]

associated with our DEA approach, it seems prudent to restrict ourselves to only a few of the variables for which data are available from the PFT experiment. This means that our application to Program Follow Through is only illustrative. On the other hand, the variables we shall study are important ones and so the adverse findings of this DEA illustration cannot be simply brushed aside. Moreover, omitted parts of the program (such as the community services components) should have biased the results in favor of PFT. In other words, even a favorable outcome for PFT would have fallen short of what is required in that further justification for these other expenditures and activities would be needed before a pro-PFT recommendation was warranted.[1] The fact that our study is not favorable to PFT compared to NFT means that strong effects in other dimensions are needed to compensate for this.

---

[1] Actually the separation between PFT and NFT is not as complete as might be desired. For one thing, other Title I experiments might have been underway in some of the NFT components and possible contaminating effects could also emerge from even social interchanges between NFT and PFT participants. See pp. 13 ff. in Vol. II-A of U.S. Office of Education [2].

Bearing the above points in mind, we now turn to another topic that also needs to be considered. Although we select and discuss the output and input variables one at a time we need to emphasize that this is not the way we shall use them. Our approach will involve uses of input and output variables considered simultaneously. Indeed, by means of the dual variables associated with our linear programming problems, we can also obtain simultaneous estimates of the relations connecting these variables to one another.[1]

In this way we avoid possible objections such as those which apply to standard statistical regression approaches that either (a) treat each of the outputs as a dependent variable in separately estimated regressions[2] or (b) scalarize all of the outputs by weighting them, e.g., relative to assumed costs and benefits, which are assigned *a priori* to secure one overall regression.[3] An easy way to summarize what we are saying is that we are simultaneously estimating all of the coefficients for an activity vector of inputs and outputs and at the same time we are using the duality relations of linear programming to effect the necessary adjustment for efficiency.[4]

---

[1] See [10] and [23] for further discussion including the possible uses of these coefficients for tradeoff analyses, etc., which we do not discuss further in the present article.

[2] See pp. 72 ff. in U.S. Office of Education Vol. II-C in [2].

[3] A use of so-called simultaneous equation (econometric) estimation methods might also be employed as in [4] and [3]. We do not use these methods here, however, because (a) they are too limited in the number of equations and variables they can handle and (b) they are not suited to estimating the *extremal* relations involved in securing the wanted efficiency ratings in any case.

[4] See the discussions in [10] and [23].

Turning now to data considerations we need to note that our analysis, and hence our results, are based on results only from Cohort II-K -- which is one of a set of two longitudinal cohorts treated by Abt Associates in [1]. [1] Our choice was dictated by a variety of considerations[2] such as the non-availability of complete Cohort III and IV[3] results at the time our research was being conducted, and by the fact that Cohort I data were available only in an aggregate form that did not permit access to the individual DMU level that is needed for the DEA procedures.

From a set of 11 output measures, three were selected for this study which we associate with their variable characterizations as follows:

y_1: Total Reading Score on the MAT = the Metropolitan Achievement Test. See [22]. This is a measure of several dimensions of a child's reading ability. The score is a site average. The test was administered on a group basis over a period of several days. As a measure of the cognitive ability of reading, it required not only word skills but also comprehension and inferential skills as well. There are clearly problems connected with the use of such "standardized reading achievement test" scores to measure the reading performance level of disadvantaged children for whom this test was admittedly not developed. However, faced with the absence of a viable alternative testing mechanism, the MAT was employed.

y_2: Total Mathematics Score on the MAT. This is a measure of a child's quantitative skill ability. This test is identical to the MAT Reading Test in its manner of administration and vulnerability to criticisms as a standardized test.[4] It attempts to measure several aspects of a child's quantitative skills make-up such as mathematics computation ability, knowledge of mathematical concepts and problem solving ability.

---

[1] The results in [2] were released only after the study we are reporting here was concluded.

[2] See [23] for further discussion.

[3] Cohort IV data are still not available.

[4] See pp. 38 ff. in U.S. Office of Education Vol. II-A [2].

y_3: Coopersmith Self-Esteem Inventory. This is a measure of a dimension of noncognitive growth or affective behavior. Developed by Stanley Coopersmith [16], the test aims at measuring aspects of a child's feelings of self-esteem. This is done via measurements of a child's feelings about himself, the way he thinks other people feel about him, and his feelings toward school. Given the stated universal Follow-Through commitment to the development of the noncognitive facets of their participants, it seems appropriate to include this as a representative indicator of affective behavior modification in our illustration.

All of these output measures were taken at the end of the third grade PFT and NFT experiences. This is the terminal or end grade for which the experiment was conducted and hence provides the cumulative effects to this point. For this illustrative exercise, it is not intended that we include all possible or available outcome measures including ones that were obtained en route to this terminus. It is felt that the three "final" output measures listed above are sufficiently representative of the others in that they include two cognitive measures and one noncognitive measure. For our purposes such a group will be sufficient in that (a) the outcomes supposedly measured by these variables are important in their own right and (b) they help to illustrate some of the difficulties that might be encountered in endeavors to weight these and other such educational outputs in any *a priori* manner.

Following these same criteria we also briefly describe our selection of 5 input variables (from a set of 25) in the following manner:[1]

x_1: Education level of mother, as measured in terms of percentage of high school graduates among the female parents. This measure was chosen because past studies indicate that it is highly correlated with home environment in its effects on academic performance. More particularly, home environment was expected to be an important input for any study of a program's effect on disadvantaged children.

x_2: Highest occupation of a family member. This again is an important non-school input. The highest occupation measure was felt to be a better indicator of social-economic status than mean income (which was also available for use in this study). To be sure, it lacks the objectivity and continuum

---

[1] See Abt [1] for a detailed description of the variables and their role in the original project.

properties of mean family income, but it provides additional information along other dimensions and does not require adjustments from nominal to real values as do income figures, e.g., in proceeding from one region of the country to another.

x_3: Parental visit index. This is a count of the number of visits to the school or with Follow Through personnel. It is supposed to be a measure of parental interest with resulting effects, especially on affective aspects of education, in these early grades.

x_4: Parental counseling index. Also called the parent-child interaction index, this is a measure of the amount of time spent by parents in interacting with the child on school related (cognitive and skill acquisition) topics -- e.g., as in reading together.

x_5: Number of teachers. This is the number of teachers at a given PFT or NFT site. This variable was intended as a simple measure of the labor intensity and/or the amount of skilled time and attention that the site (school) was willing to devote to the program.

Regarding the actual recording methods employed for the data on these variables we here note that for each site we obtain a vector of input and output variable observations.[1] For that point in time (the same time measurement point for all sites) these measurements represent the total or cumulative (not average) level of output performance or total level of input magnitude over the site.[2] Tables A-3 and A-4 in the Appendix give the output data for PFT and NFT, respectively, while Tables A-5 and A-6 contain the input data for PFT and NFT with respect to the variables we are using.

---

[1] Data quality and completeness, along with methods of collection and validation are discussed in [1] and [2]. See also pp. 42 ff. in U.S. Office of Education Vol. II-A [2].

[2] The total magnitudes represent the mean values of outputs and inputs multiplied by the total number of students per site, where the measurement unit per student is hundred students. Note there is one exception to this routine, namely, x_5, number of teachers, where the original study information was already provided in total site magnitude form. Tables A-1 through A-4 list the total measurements on the three outputs and five input variables for both PFT and NFT.

6. DMU AND MANAGEMENT DECISION-MAKING EFFICIENCY

Other approaches to efficiency evaluation in education have also been conducted in ways that are related to ours.[1] These have been concerned only with estimating or detecting inefficiencies in a collection of DMU's or with identifying organizations or programs where such inefficiencies are present. In our case, however, we want to disentangle managerial (= decision making) efficiency from program efficiency, if possible, before reaching a judgement on the latter. Naturally we do not expect a mechanical procedure such as DEA to supply all of the answers as to the nature and sources of such inefficiencies. It should at least supply guidance, however, so that by audit follow-up or other such on-the-site study techniques one can obtain further confirmation and perhaps even specific remedies or correctives.[2] Thus, we do not regard our analysis as causal, at least for the purposes being considered in this paper. I.e., we are not effecting imputations in the sense that one may perhaps assign causal significance to the independent variable in a regression model approach.[3] We seek only initial efficiency evaluations which are sufficiently well grounded so that (a) other evidence is needed to prevail against these evaluations and/or (b) guidance as to where to look for possible

---

[1] See, e.g., [7], [19] and [24].

[2] This may be thought of in terms of a rough analogy to statistical quality control procedures where observations or estimates that exceed control limits and/or the presence of runs in sequences of observations provides guidance as to when and where to look for trouble and possible remedies.

[3] See, e.g., H.A. Simon [27].

improvements is provided in the senses that were just indicated.

Reference to Table 1 shows that we have 49 DMU's for PFT and 21 for NFT. In each case we first effect our efficiency evaluations relative to members of the same set. That is, we distinguish an $\alpha=1$ set for the DMU's in PFT and an $\alpha=2$ set for the DMU's in NFT. In each case the efficient members of the set generate an efficiency frontier, or "envelope," which represents the boundary of the production possibilities (as indicated by the observations) in each case. See the discussion at the end of the next section.

We shall refer to these as "α-envelopes" which we may also specialize to an "α=1-envelope" and an "α=2-envelope" and, on occasion, we will replace these by references to the PFT and NFT envelopes, respectively. After these envelopes have been derived we shall then generate a further envelope that we shall refer to as an "inter-program envelope" or, more briefly, an "inter-envelope." This latter will then be used as a reference for judging the "program efficiencies" of PFT and NFT after first bringing all DMU's onto their respective α-envelopes (in order to eliminate inefficiencies resulting from managerial inefficiencies in the observation for any DMU.)

To initiate this process we refer to Table 2, below, which shows the results of the program specific applications of (2) for $\alpha=1$ (PFT) and $\alpha=2$ (NFT), respectively. We refer to these as "managerial" efficiency calculations on the supposition that these DMU's are referenced only to others that share the same program constraints. Hence the observed variations are ascribed to variations in individual DMU (managerial) decision-making[1] within each of PFT or NFT. In Table 2 the "$h_o^{*\alpha}$" values which are equal to "1" are for DMU's which are on the efficiency frontier, i.e., on the "α-envelope." All

[1] Including technological and organizational choices perhaps imposed by past managers and hence beyond current managerial control but which nevertheless affect the way present decisions are made.

Table 2

PFT and NFT Program Specific α-Envelope Efficiency Values

PFT Site #	h_0^{*1} Efficiency Value	NFT Site #	h_0^{*2} Efficiency Value
1*	1.00	50	0.95
2	0.90	51	0.92
3	0.98	52*	1.00
4	0.90	53	0.87
5*	1.00	54*	1.00
6	0.90	55*	1.00
7	0.89	56*	1.00
8	0.91	57	0.92
9	0.87	58*	1.00
10*	1.00	59	0.92
11	0.98	60	0.98
12	0.97	61	0.88
13	0.86	62*	1.00
14	0.98	63	0.96
15*	1.00	64	0.91
16	0.95	65	0.97
17*	1.00	66	0.92
18*	1.00	68*	1.00
19	0.95	69*	1.00
20*	1.00	70	0.94
21*	1.00		
22*	1.00		
23	0.96		
24*	1.00		
25	0.97		
26	0.93		
27*	1.00		
28	0.94		
29	0.84		
30	0.90		
31	0.83		
32	0.90		
33	0.94		
34	0.85		
35*	1.00		
36	0.80		
37	0.94		
38	0.94		
39	0.91		
40*	1.00		
41	0.94		
42	0.94		
43	0.87		
44*	1.00		
45	0.89		
46	0.90		
47*	1.00		
48*	1.00		
49*	1.00		

*** Denotes a site with an efficiency value of "1"**

other "$h_o^{*\alpha}$" < 1 are contained within the boundaries of their respective " α- envelope" and are thus less efficient.

[Insert Table 2]

Having obtained these program specific values for our measure of "managerial efficiency," we next examine the results to ascertain whether these efficiencies differ between the two sets. Toward this end, we have at our disposal a number of classical statistical tests and other measures. By way of illustration, we choose two such comparisons. The first is a comparison of differences in the probability of a PFT versus an NFT site being on their respective "α-envelopes." In other words, we seek a comparison of the probability of a PFT versus an NFT site having an efficiency value of 1.0 relative to its own α-envelope. Without attempting to exhaust the possiblities of simple statistical measures that our approach permits, we shall then proceed to our second statistical test and average these managerial (= DMU) efficiencies by reference to their respective $h_o^{*\alpha}$ values. In this way we shall be able to allow for "magnitude" as well as "probability" in the simple comparisons that we shall effect in order to ascertain whether we get similar or different results from these two approaches.

Our first measurement is addressed to the probabilistic question that we have just now formulated. To treat this question, we simply consider the ratio of the number of DMU's from a given α reference set which are on their respective envelopes -- i.e., have $h_o^{*\alpha}$ values of 1.0 -- relative to the total number of DMU's in this same feference set. Using the information from Table 2, we therefore find that the probability of being on the α = "1-envelope" is

$$P(\alpha=1) = \frac{m_e(1)}{m(1)} = \frac{17}{49} = \frac{51}{147}$$

where $m_e(1)$ is the number of PFT DMU's with h_o^{*1} values of 1.0 and m(1) is the number of DMU's in PFT. For NFT this probability is

$$P(\alpha=2) = \frac{m_e(2)}{m(2)} = \frac{8}{21} = \frac{56}{147}$$

where m(2) refers to the total number of DMU's in NFT and $m_e(2)$ refers to the number on the efficiency frontier.

Proceeding now in a somewhat informal manner, we can see that

the managers in the two programs, PFT and NFT, have about an equal likelihood of being on their respective program referenced efficiency frontiers.

The differences between $P(\alpha=1)$ and $P(\alpha=2)$ are not statistically significant so that an implication of the above results is that both PFT and NFT site managers or "decision-makers" are drawn from the same managerial efficiency pool or population. Relatively speaking there appear to be just as many managers in PFT as in NFT who, within the limits of their program constraints, operate their sites efficiently or inefficiently.

The above probability measures provide insight only into the location of DMU's in PFT and NFT relative to their respective boundaries. Questions involving differences between the distribution of efficiency values in the two groups are not addressed by the above calculation. Thus to check and perhaps extend our understanding of these results, while still staying with relatively simple measures, we consider a comparison between the mean efficiency value differences in the two sets. Again using the information contained in Table 2, we compute a PFT mean efficiency as $\bar{h}_o^{*1} = 0.946$ and an NFT mean efficiency value as $\bar{h}_o^{*2} = 0.958$. Thus, NFT has a slightly higher mean efficiency.

Proceeding now in a more formal manner than before we proceed to a two-tailed unpaired t-test comparison of the significance of these differences between mean efficiency values. For this purpose, we assume statistical independence between observations, but <u>do not</u> assume that the two population variances are the same. Hence, recourse is to the so-called Behrens-Fisher statistic [1] for which we obtain

$$\hat{t}_{1,2} = -0.26603$$

---

[1] We actually used Cochran's approximation formula as given on p. 115 of [25].

by reference to data of Table 2. At the 0.05 level this is not significant and so we do not reject the null hypothesis that PFT and NFT pools do not differ in their managerial efficiency.

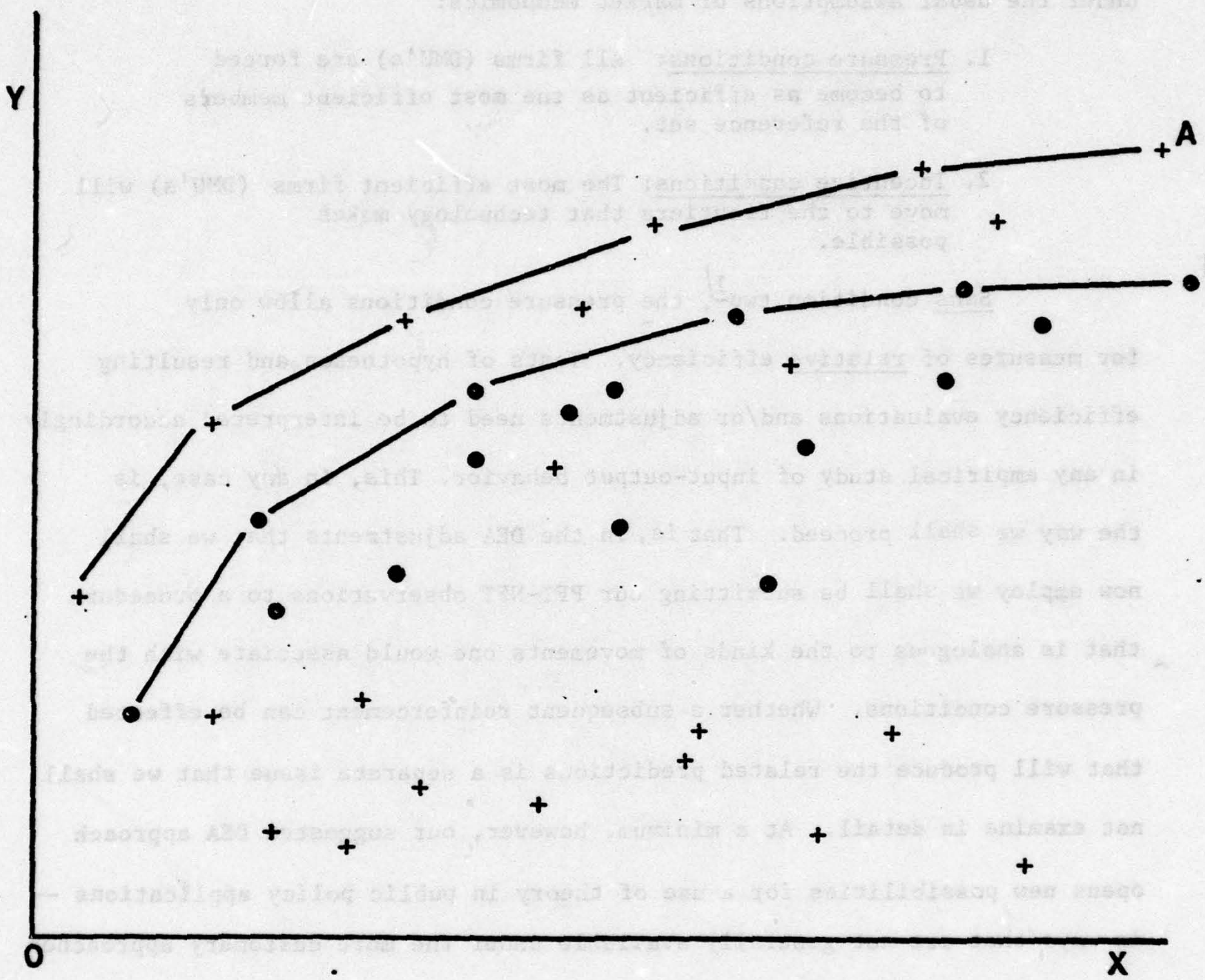
Before proceeding to our wanted "program efficiency" comparisons we pause to try to illuminate some of the assumptions (and possibilities) underlying our approach as follows. First we consider the hypothetical observations portrayed in Figure 1. Here we are supposing two sets of DMU's which have similar outputs and inputs. All outputs and inputs are fixed at the same level for every DMU except for the one input in amounts, x, and the one output in amounts, y, shown in Figure 1.

If one were simply to calculate average efficiencies, the set A would rank lower than B. On the other hand, the efficiency frontier for A dominates that of B so that the simple calculation of these averages might lead to erroneous inferences concerning the efficacy of B vs. A. At a minimum, therefore, one ought to try to detect the presence of different sources of inefficiency before assigning them all to the programs associated with A and B, respectively.

We shall shortly be adjusting all points up to the frontiers to effect our comparisons and, of course, the kinds of predictions that one might make from this quarter are different than those one might make by effecting ordinary statistical estimates from the data of Figure 1. In particular one will now need to effect supplementary analyses and perhaps provide guides and/or controls for the managers associated with the DMU's in such sets as A and B in order to reenforce the predictions that are being made under our DEA approaches.

In terms of ordinary private sector (market) economies one may think of the A and B frontiers in the manner of "technologies" that limit what is possible. If one were to suppose a free play of competition

Figure 1



Legend:

Pluses(+) = Observations for DMU's in A

Dots (●) = Observations for DMU's in B

then one might also suppose that the following conditions are also at work under the usual assumptions of market economics:

1. Pressure conditions: All firms (DMU's) are forced to become as efficient as the most efficient members of the reference set.

2. Incentive conditions: The most efficient firms (DMU's) will move to the frontiers that technology makes possible.

Sans condition two[1], the pressure conditions allow only for measures of relative efficiency. Tests of hypotheses and resulting efficiency evaluations and/or adjustments need to be interpreted accordingly in any empirical study of input-output behavior. This, in any case, is the way we shall proceed. That is, in the DEA adjustments that we shall now employ we shall be submitting our PFT-NFT observations to a procedure that is analogous to the kinds of movements one would associate with the pressure conditions. Whether a subsequent reinforcement can be effected that will produce the related predictions is a separate issue that we shall not examine in detail. At a minimum, however, our suggested DEA approach opens new possibilities for a use of theory in public policy applications -- in ways that are not generally available under the more customary approaches that have heretofore been used.

---

[1] Actually it is not clear how the validity of the assumptions associated with these incentive conditions can be tested unless one has access to other information (e.g., engineering information) besides market observations.

7. THE INTER ENVELOPE AND PROGRAM EFFICIENCY

Reference to Figure 2 may help to show how we propose to effect our "program efficiency" comparisons. In this diagram the dots (·) and the x's are supposed to represent PFT and NFT observations, respectively. These are all hypothetical observations arranged to suit our convenience. The resulting two dimensional portrayal is intended to show the amounts of two inputs required to produce one unit of the same output by each of several different DMU's in PFT and NFT, repectively.

The α-envelopes for PFT and NFT now correspond to the so-called "unit isoquant."[1] These envelopes (or isoquants) are determined by applying the linear programming equivalent of (2) to the observations for $\alpha = 1$ and $\alpha = 2$ in turn. Points such as A represent PFT observations which have values $h_o^{*1} < 1$ since these are not on the PFT α-envelope for which the values $h_o^{*1} = 1$ apply. Similarly, points such as B and C have values of $h_o^{*2} < 1$ determined relative to the NFT -envelope.

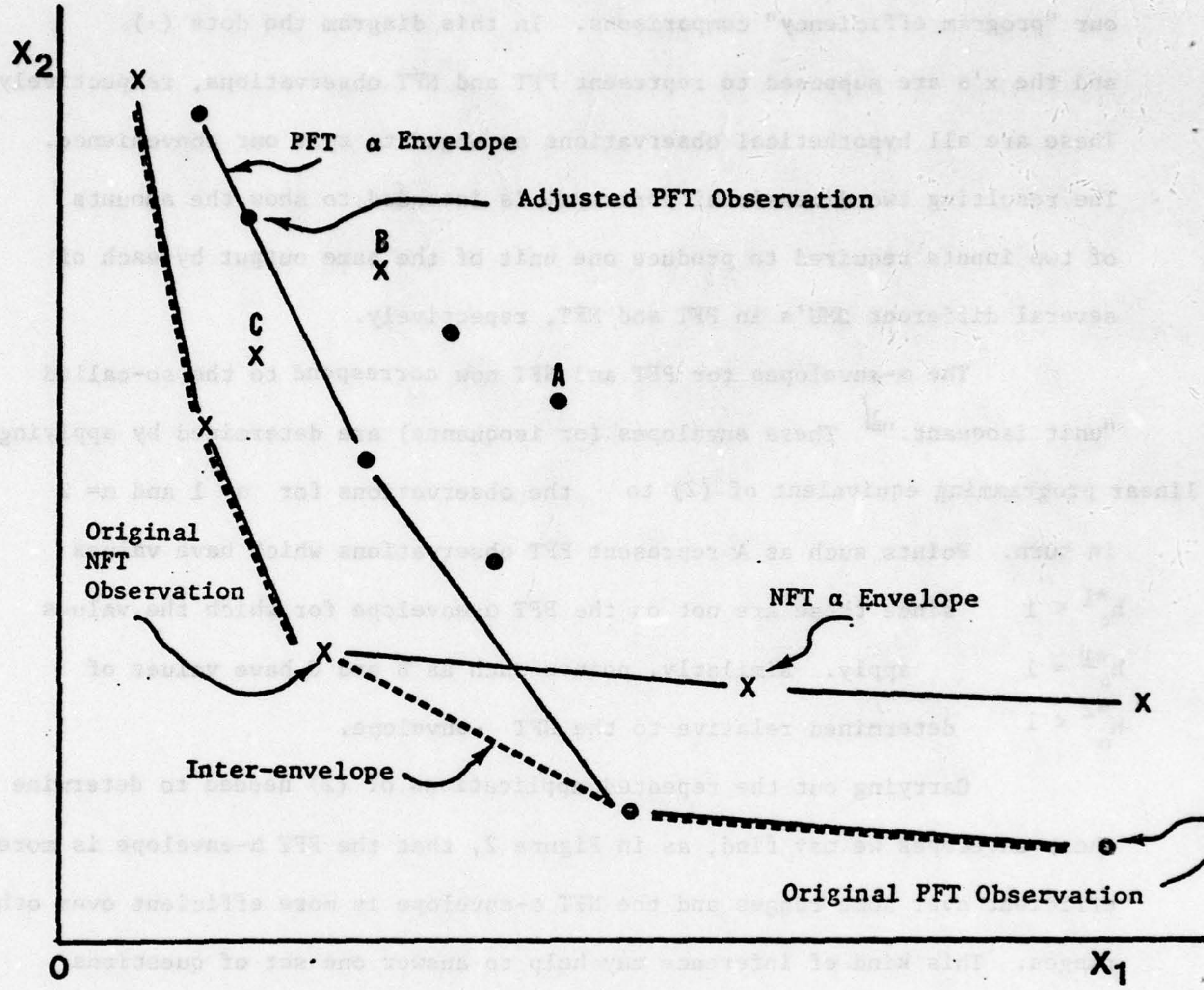
Carrying out the repeated applications of (2) needed to determine these envelopes we may find, as in Figure 2, that the PFT α-envelope is more efficient over some ranges and the NFT α-envelope is more efficient over other ranges. This kind of inference may help to answer one set of questions but we are now seeking an evaluation that will help us to decide which program is better in some overall efficiency sense.

To help in answering the latter question we proceed as follows.

---

[1] See the discussion in section 1.

Figure 2



Legend:

● = PFT Observations

X = NFT Observations

●—● = α Envelope for PFT

X—X = α Envelope for NFT

------ = Inter-envelope

First we utilize the procedures described in [10] to bring all of the observations onto their respective α-envelopes. Then we construct an inter-envelope that will enable us to compare the resulting clusters of DMU's on the assumption that they are all operating on the efficiency boundaries permitted by their program constraints. This permits us to impute remaining differences to the respective programs by reference to a common envelope which is always at least as efficient as any of the α -envelopes -- as witness the "inter-envelope" portrayed in Figure 2.[1]

To make the above instructions more concrete we now replace (2) with the following formulation for effecting the inter-envelope efficiency determinations.

$$\max.\ h_o = \frac{\sum_{r=1}^{s} u_r \hat{y}_{ro}}{\sum_{i=1}^{m} v_i \hat{x}_{io}}$$

subject to

$$(3) \qquad 1 \geq \frac{\sum_{r=1}^{s} u_r \hat{y}^1_{rj}}{\sum_{i=1}^{m} v_i \hat{x}^1_{ij}}\ ; \quad j=1, \ldots, m_1$$

and

$$1 \geq \frac{\sum_{r=1}^{s} u_r \hat{y}^2_{rj}}{\sum_{i=1}^{m} v_i \hat{x}^2_{i}}\ ; \quad j=1, \ldots, m_2$$

where, as before, all variable values are constrained to be positive. Here the caret over a letter indicates that the efficiency adjustments to the envelopes for α=1 and α=2 have been carried out as described in the preceding paragraph.

1 For ease of understanding we are conducting this portion of the discussion as though the isoquant assumption is valid. However, we are dealing with multiple outputs and inputs so that the concept of an isoquant has no meaning and must be replaced by the more general concept of a production possibility set. See [10]. Also, we shall continue (in this same spirit) as though we are concerned with resource conservation possibilities. Actually many of our inputs are fixed, outside the realm of managerial discretion, and so one might more properly speak of output augmentations rather than input reductions along the lines that are also indicated in [10]. See also the discussion following (1) in section 1.

The DMU_o being rated in (3) can come from either $\alpha = 1$ or $\alpha = 2$. The fact that a DMU is efficient under either of these programs, however, does not necessarily produce an $h_o^* = 1$ for this DMU. Failure to achieve this rating, however, is assumed to be due to the program constraint under which this DMU was operating when the adjustments from y_{ro} and x_{ij} to $\hat{y}_{ro}$ and $\hat{x}_{ij}$ were effected. In other words, an $h_o^* < 1$ is now attributed to the program rather than the DMU being evaluated in each case.

6. AN INFORMATION THEORETIC TEST OF PROGRAM EFFICIENCY

The indicated application of (3) yields the h_o^* values shown in Table 3. We could, of course, now repeat the same kind of analysis that we undertook for effecting efficiency comparison between DMU's operating under PFT and NFT, respectively. Something more is wanted, however, in that our program comparisons should really be effected relative to the clusters of managers operating under each program. In short, we would like to effect our comparison by means of a measure of the distance between the two distributions exhibited in Table 3.

A variety of measures of this kind are available. The one we select involves an extended form of the information statistic which Kullback [19] refers to as a measure of "divergence."[1] This measure is defined as

(4) $$J(f_1, f_2) = I(f_1 : f_2) + I(f_2 : f_1)$$

where f_1 and f_2 represent density functions, discrete or otherwise, and

---

[1] See Kullback [21] p.190 for a discussion of the relation of this measure to Mahalanobis' "D^2" or "generalized distance measure."

I represents the information statistics given by

$$I\ (f_1: f_2) = \int f_1\ (x)\ \log \frac{f_1\ (x)}{f_2\ (x)}\ dx$$

(4.1)

$$I\ (f_1: f_2) = \int f_2\ (x)\ \log \frac{f_2\ (x)}{f_1\ (x)}\ dx$$

when the densities are continuous. Inserting these two expressions in (4) produces

(4.2) $$J\ (f_1, f_2) = \int (f_1\ (x) - f_2(x))\ \log \frac{f_1(x)}{f_2(x)}\ dx$$

which shows that J is a measure of the "difference" between these two densities.

Actually $J\ (f_1, f_2)$ has most of the properties of a distance function.[1] Each of the terms in the sum defining (4) is nonnegative and so,

(5.1) $J(f_1, f_2) \geq 0$, with $J\ (f_1, f_2) = 0$ if and only $f_1 = f_2$

which is, as we know, the nonnegativity requirement plus positivity for distinct points for any metric distance function. Similarly, this divergence as defined above has the symmetry property, <u>viz.</u>,

(5.2) $J\ (f_1, f_2) = J\ (f_2, f_1)$.

The only property that J lacks is the so-called "triangle property," <u>viz.</u>, we cannot guarantee that

(5.3) $J(f_1, f_3) \leq J(f_1, f_2) + J(f_2, f_3)$.

---

[1] See Appendix A in Charnes and Cooper [8] , pp. 154-156 for a treatment of the three defining properties for a metric distance function.

That is, we may have

(5.4) $\quad J(f_1, f_3) > J(f_1, f_2) + J(f_2, f_3).$

Statistically speaking this means that the divergence between f_1 and f_3 may be significant even though the sum of the divergence between (f_1, f_2) and (f_2, f_3) is not statistically significant. In other words, we cannot rely on the results of the latter to test the former.

Actually, we do not need the latter for our purposes. Ours is not even a symmetrical comparison, in that we are only testing whether PFT is significantly better than NFT relative to the "inter" frontier. That is, we consider both PFT and NFT estimated program efficency values with respect to the hypothesized frontier values of $h_o^{*\alpha} = 1$. Toward this end we need only apply what Kullback refers to as the "directed divergence" -- viz., $I(f_1:f_2)$ is one such directed divergence in (4.1) and $I(f_2:f_1)$ is the other. The former, we may say, measures the divergence between f_1 and f_2 from the standpoint of f_1 and the latter measures it from the standpoint of f_2. Of course, these measures will not, in general, be the same.[1]

In our case we are using the "inter" envelope as the common standard of reference. That is, we are not comparing the α-envelope distributions directly but are, instead, comparing each of them to the inter-envelope distribution. Thus, using the data contained in Table 3.

---

[1] For other references that involve new and somewhat surprising uses of these so-called "Kullback-Liebler" statistics see [9] and [25]. See especially the further references to the important articles by Akaike that are also cited in [9] and [25]. See also [13] for a further extension that discusses recently developed duality relations for constrained Khinchin-Kullback-Leibler estimates.

Table 3

Inter-Envelope Efficiency Values

PFT Site #	h_0^* Efficiency Value	NFT Site #	h_0^* Efficiency Value
1	0.92	50*	1.00
2*	1.00	51*	1.00
3	0.94	52*	1.00
4*	1.00	53*	1.00
5	0.93	54*	1.00
6*	1.00	55	0.99
7	0.99	56*	1.00
8*	1.00	57*	1.00
9	0.98	58*	1.00
10	0.92	59*	1.00
11*	1.00	60	1.00
12*	1.00	61*	1.00
13	0.99	62*	1.00
14	0.95	63*	1.00
15*	1.00	64*	1.00
16*	1.00	65*	1.00
17*	1.00	66*	1.00
18*	1.00	67*	1.00
19	0.99	68	0.99
20*	1.00	69*	1.00
21*	1.00	70*	1.00
22*	1.00		
23	0.99		
24*	1.00		
25*	1.00		
26	0.99		
27*	1.00		
28*	1.00		
29	0.99		
30*	1.00		
31	0.99		
32*	1.00		
33	0.99		
34	0.98		
35*	1.00		
36*	1.00		
37	0.94		
38	0.99		
39*	1.00		
40	0.95		
41	0.99		
42*	1.00		
43	0.99		
44*	1.00		
45	0.99		
46*	1.00		
47*	1.00		
48*	1.00		
49*	1.00		

*Denotes a site with an efficiency value of "1"

we obtain

$$I\ (f_1: \delta) = 2.40226 \times 10^{-4}$$

(6) and

$$I\ (f_2: \delta) = 0.03684 \times 10^{-4}$$

where $f_1 = f_1\ (h_o^*)$ and $f_2 = f_2\ (h_o^*)$ represent the distributions[1] portrayed for PFT and NFT, respectively, and

(7) $$\delta\ (h_o^*) \text{ gives } \begin{cases} \delta = 1 \text{ for } h_o^* = 1 \\ \delta = 0 \text{ for } h_o^* < 1 . \end{cases}$$

Our comparison is with the <u>degenerate</u> distribution for which all $h_o^*=1$ so that also $\delta(h_o^*) = 1$. In this case directed divergence, therefore, reduces to the entropy measure of "disorder," or, alternatively, "divergence" from the situation in which all $h_o^*=1$ -- which, we may note, is a consequence of the assumptions discussed in connection with the pressure and/or incentive conditions noted at the close of the last section. Thus, referring to the directed divergences exhibited in (6), in accordance with our just noted assumptions, we observe that the first statistic in (6) has an I value for f_1 which exceeds the I value for f_2 and hence has a greater divergence value than f_2. No significance test will reverse this sign difference and hence we now conclude that our evidence is to the effect that PFT has not demonstrated its superior efficiency. This is consistent with our preceding results, too, and hence for reasons advanced earlier (e.g., the additional expenditures involved) an implementation of PFT is not warranted from this evidence.[3]

[1] See [11] for a detailed discussion of these kinds of statistical distributions.

[2] The statistic for I. which is the so-called "Kullback-Leibler" statistic is asymptomatically distributed as χ^2 under a reasonably broad class of conditions. Hence recourse may be had to this property when significance tests are wanted. See Kullback [21].

[3] In a general way this agrees with the findings -- but without the numerous qualifications and exceptions -- described on pp. 158 ff. of U.S. Office of Education Vol. IIA [2].

Of course, our PFT-NFT evaluations need not (and should not) end here. A variety of additional possibilities are also open for study. We might, for instance, conduct a facet-by-facet comparison of the DMU's on each α-envelope.[1] Note, for example, that DMU's on the same facet may be identified explicitly by the fact that they will have the same optimal bases. Furthermore, the direction numbers (and hence the direction cosines)[2] for determining the distance of these DMU's from the relevant part of the inter-envelope are also at hand. Hence, measures of average distance from the inter-envelope can be readily secured and applied facet by facet, if desired, for further evaluation of subsets of PFT-NFT possibilities.

---

[1] See, e.g., Gray [17].

[2] See Appendix A in [7] which shows how to obtain the distance from a point to a hyperplane by means of these values.

Returning to Figure 2 we may also observe that still additional possibilities are present. Note, in particular, the broken line portion of the inter-envelope which lies below <u>both</u> α-envelopes. This, we should say, arises from possible combinations of elements from PFT and NFT which offer greater efficiency possibilities than either of them within the input region subtended by this facet. Such new possibilities also need to be confirmed by further study, preferably in the field, but the point is that they should not be discarded simply because the original design did not explicitly consider these potential combinations of PFT and NFT elements.

SUMMARY AND CONCLUSION

The points that have just been made should suffice to indicate some of the possibilities that our DEA approach may offer. Here we have presented this approach in terms of an illustrative application to

Program Follow Through. It is not to be regarded as limited to this Program, however, or even to education programs. Our intention is to provide a general set of concepts and methods that can be applied to a variety of public programs where profit/cost and like considerations are not directly applicable.[1] The point to bear in mind is that these concepts are at their best when applied to situations in which there are an agreed upon set of objectives and in which resource diversions to other programs are not at issue.[2]

Where these conditions are met there is still an interest in resource conservation, on the presumption that released resources are of use elsewhere. Notice now that our DEA approach gives us a method of ascertaining the <u>amount</u> of resource conservation and/or output augmentation that is possible as well as a way of distributing these amounts between program and managerial efficiency. How any of the conserved amounts of resources might best be redistributed to other activities, e.g., to activities of a non-education variety, involves issues of pricing and weighting that are not addressed in our formulations.

Another point of interest is our choice of the Kullback-Leibler statistic. This choice was elected because we wanted to be able to compare the distributions of f_1 and f_2 with the further possibilities that might be available by relaxing the program boundaries. This could not

---

[1] See [10] for further discussion.

[2] In the terminology of the U.S. General Accounting Office we are here concerned with efficiency (including economy) and not "effectiveness" and/or "propriety." See [14] and [31].

have been done as easily or directly by comparing f_1 and f_2, via, e.g., $I(f_1, f_2)$, although, of course, such a comparison might have been effected and even extended for other purposes by recourse to (4) ff.

It should perhaps be again noted that we have here reversed the usual relation between statistical methods and economic theory in empirical research. A great deal of the latter research has been directed toward theory testing, of course, and that is not our objective here. We are instead concerned with using that theory (e.g., accepted parts of production theory) to assist in the evaluation of public policy programs. In addition to arriving at evaluations for these programs (and their management) we have also been concerned with using theory to uncover opportunities for resource conservation or output improvements that would otherwise remain hidden from view.

Of course, we have ~~also~~ confined our modeling of possible new opportunities by reference to inferences from observed data. We have also suggested that our DEA approach is best regarded as a guide and that it requires supplementation by further study, preferably in the field, in order to ensure that the indicated opportunities are really present. The fact of their presence is also not decisive unless controls or other alterations can be specified (e.g., by program audits of GAO variety)[1] to ensure that the indicated improvement possibilities will be forthcoming .

---

[1] Or by suitably extended versions of such audits to allow for inter-program comparisons and evaluations. See Churchill, et.al. [14].

In conclusion we might again contrast our DEA analysis as an example of "prediction under control" in comparison with the "pure prediction" that is represented by the following statement by Milton Friedman:[1]

> The only relevant test of the validity of a hypothesis is comparison of its predictions with experience.

Evidently Professor Friedman thinks that the end of theory is prediction in the uncontrolled sense. This is one valid view of scientific research, to be sure, but it is not the only one. At a still deeper level one may say that a scientific theory achieves an even greater value when it tells us where to look for possible modes of behavior that might otherwise be missed entirely.

---

[1] From [18], pp. 8-9.

BIBLIOGRAPHY

[1] Abt Associates, Education as Experimentation: A Planned Variation Model, Vol IA and IB (Cambridge, Mass: Abt Associates, Inc., 1974).

[2] _________, Education as Experimentation: A Planned Variation Model, Vol, IIIA and IIIB (Cambridge, Mass: Abt Associates, Inc. 1976).

Also issued as The Follow Through Planned Variation Experiment, Vols. II A, II B and II C (Washington: U.S. Office of Education, 1977).

[3] Boardman, A. "Policy Models for the Management of Student Achievement and Other Education Outputs," TIMS Studies in the Management Sciences, Vol. 8, Management Science Approaches to Manpower Planning and Organization Design (Amsterdam's North Holland Publishing Co., 1978).

[4] _________, O.A. Davis and P.R. Sanday, "A Simultaneous Equations Model of the Educational Process," Journal of Public Economics, 7, 1977, pp.23-49.

[5] _________ and R. J. Murnane, "The Use of Panel Data in Education Research," Sociology of Education (forthcoming).

[6] Campbell, D.T. and J.C. Stanley. Experimental and Quasi-Experimental Designs for Research, (Chicago: Rand-McNally College Publishing Co.,1963).

[7] Carlson, D.E. The Production and Cost Behavior of Higher Education Institutions (Berkeley, Calif: University of California, 1972).

[8] Charnes, A. and W.W. Cooper. Management Models and Industrial Application of Linear Programming, Vol. I (New York: John Wiley and Sons, Inc., 1961).

[9] _________, _________ and D. B. Learner, "Constrained Information Theoretic Characterizations in Consumer Purchase Behavior," Journal of the Operational Research Society 29, No. 9, 1978, pp. 833-884.

[10] _________, _________, and E. L. Rhodes, "Measuring the Efficiency of Decision-Making Units," European Journal of Operational Research, 2, No. 6, Nov. 1978, pp.429-444. See also "Corrections," op. cit.,forthcoming.

[11] _________, _________, and _________, "On the Distribution of Efficiency Measures for Decision Making Units," University of Texas Center for Cybernetic Studies, Research Memorandum, Austin, Texas, Feb., 1978.

[12] _________, _________ and A. Schinnar, "A Bi-Extremal Principle for Estimating Extremal Relations from Empirical Data." Research Report, University of Texas, Center for Cybernetic Studies (Austin: University of Texas at Austin, 1979).

[13] __________, __________, and L. Seiford, "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leibler Estimation," Math Operationsforsch. Statist. Ser. Optimization, 9, No. 1, 1978, pp.21-29.

[14] Churchill, N.C., W. W. Cooper, J.San Miguel and V. Govindarajan and J. Pond, "Comprehensive Audits: Some Findings and Some Suggestions for Research," Symposium on Auditing Research, II (Department of Accounting, University of Illinois at Urbana-Champaign, 1977).

[15] Cooper, W.W., "Understanding Prediction and Control -- And Other Matters Relating to Scientific Research," SUPALUM Alumni Magazine, School of Urban and Public Affairs, Carnegie-Mellon University, No. 2, April, 1974, pp.25-28.

[16] Coopersmith, S. The Antecedents of Self Esteem, San Francisco

[17] Farrell, M. J. "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society 120, Sec. A, pt. 3, 1957, pp. 253-290.

[18] Friedman, M. Essays in Positive Economics,(Chicago: University of Chicago Press, 1953).

[19] Gray, R. and Weldon, K. , "An Experiment with Convex Production Functions" Working Paper. National Center for Higher Education Management Systems, (Boulder, Colo.: Nov. 1978).

[20] Griliches, Z., "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica, XLV, No. 1 (Jan., 1977), pp.1-22.

[21] Kullback, S. Information Theory and Statistics (New York: Dover Publications, Inc. 1968).

[22] Metropolitan Achievement Tests Special Report (New York: Harcourt Brace Jovanovich, 1971), pp. 9-11.

[23] Rhodes, E. L. Data Envelopment Analysis and Related Approaches for Measuring the Efficiency of Decision-Making Units with an Application to Program Follow Through in U.S. Education, unpublished Ph.D. thesis, (Pittsburgh: Carnegie-Mellon University School of Urban and Public Affairs, 1978).

[24] Rodgers, K. W., "The Realization of National Policy Objectives by Historically Black Colleges," (Cambridge, Mass: Arthur D. Little, Inc. 1976).

[25] Sawa, T., "Information Criteria for Discriminating Among Alternative Regression Models," Econometrica 46, No. 6, Nov. 1978, pp. 1273-1291.

[26] Shephard, R.W., Cost and Production Functions, (Princeton: Princeton University Press, 1953).

[27] Simon, H. A., "Spurious Correlation: A Causal Interpretation", in H. M. Blalock, Jr., ed. Causal Models in the Social Sciences (Chicago: Aldine Publ. Co., 1971).

[28] Snedecor, G. W. and W. G. Cochran. Statistical Methods, 6th edition, (Ames, Iowa: The Iowa State University Press, 1967). p.115.

[29] Stallings, Jane A. "Follow Through Program Classroom Observation Evaluation, 1971-72, (Menlo Park, Calif. Stanford Research Institute 1973).

[30] Summers, A. A. and B. L. Wolfe, "Which School Resources Help Learning? Efficiency and Equity in Philadelphia Public Schools," Business Review of Federal Reserve Bank of Philadelphia, Feb., 1975. See also Summers and Wolfe "Do Schools Make a Difference?" American Economic Review (forthcoming).

[31] U.S. General Accounting Office, Follow Through: Lessons Learned From Its Evaluation and Need to Improve Its Administration, (Washington: U.S. General Accounting Office, 1975).

Table A-1

FOLLOW THROUGH APPROACHES AND ASSOCIATED SPONSORS INCLUDED IN DATA ENVELOPMENT ANALYSIS STUDY

Approach and Sponsor	Number of PFT Sites	Number of NFT Sites
RESPONSIVE EDUCATION PROGRAM Far West Laboratory for Educational Research and Development	7	7
TUSCON EARLY EDUCATION MODEL (TEEM) Arizona Center for Early Childhood Education	4	2
BANK STREET COLLEGE OF EDUCATION APPROACH Bank Street College of Education	5	3
DIRECT INSTRUCTION MODEL (DIM) University of Oregou -College of Education	5	3
BEHAVIOR ANALYSIS APPROACH (BA) Support and Development Center for Follow Through - University of Kansas	6	1
COGNITIVELY ORIENTED CURRICULUM MODEL High/Scope Educational Research Foundation	3	-
FLORIDA PARENT EDUCATION MODEL University of Florida	4	3
EDC OPEN EDUCATION FOLLOW THROUGH PROGRAM Education Development Center	2	-
SELF-SPONSORED - New York City,NY	1	-
SELF-SPONSORED - Philadelphia , PA	1	1
SELF-SPONSORED - Detroit, MICH	1	-
SELF-SPONSORED - Portland,OR	1	-
SELF-SPONSORED - San Diego, CA	1	-
INTERDEPENDENT LEARNING MODEL (ILM) New York University - Institute for Developmental Studies	1	-
LANGUAGE DEVELOPMENT (BILINGUAL) EDUCATION APPROACH Southwest Educational Development Laboratory (SEDL)	2	1
HOME-SCHOOL PARTNERSHIP: A MOTIVATIONAL APPROACH Southern University and A&M College	1	-
CALIFORNIA PROCESS MODEL California State Department of Education-Division of Compensatory Education	4	-
Total Number of Sites	49	21

Source: Abt [2], p.A-18

Table A-2

Site Level Distribution of DEA Study Sample

PFT Site #	NFT Site #	Model and Site Name	Region[1]	City Size[2]	PFT[3] Student Pop.	NFT[4] Student Pop.
		Responsive Education Model				
1	30	Berkeley,CA	W	Medium City	99	71
2	51	Buffalo,NY	NE	Large City	77	27
3	52	Duluth,MN	NC	Medium City	77	79
4	53	Fresno,CA	W	Medium City	48	54
5	54	Lebanon,NH	NE	Rural Area	14	97
6	55	Salt Lake,UT	W	Medium City	36	51
7	56	Tacoma,WA	W	Medium City	51	42
		TEEM Model				
8		Baltimore,MD	S	Large City	99	-
9		Lakewood,NJ	NE	Small City	80	-
10	57	Lincoln,NB	NC	Medium City	96	55
11	58	Wichita,KS	NC	Large City	84	36
		Bank Street Model				
12	59	New York,NY	NE	Large City	72	245*
13	60	Philadelphia,PA	NE	Large City	80	37
14		Brattleboro,VT	NE	Small City	20	-
15	61	Fall River,MA	NE	Medium City	39	18
16		Wilmington,DE	S	Medium City	109	
		DIM Model				
17		New York,NY	NE	Large City	31	-
18	62	E.St. Louis,IL	NC	Large City	56	21
19		Grand Rapids,MI	NC	Medium City	103	
20	63	Racine,WI	NC	Medium City	62	27
21	64	Flint,MI	NC	Medium City	77	66
		BA Model				
22		New York,NY	NE	Large City	43	-
23	65	Philadelphia,PA	NE	Large City	108	27
24		Portageville,MO	NC	Rural Area	47	-
25		Kansas City,MO	NC	Large City	61	-
26		Louisville,KY	S	Large City	90	-
27		Meridian,IL	NC	Rural Area	68	-
		Cognitive Curriculum Model				
28		New York,NY	NE	Large City	52	-
29		Chicago,IL	NC	Large City	18	-
30		Okaloosa Co.,FL	S	Small City	48	-
		Parent Education Model				
31		Philadelphia,PA	NE	Large City	46	-
32	66	Jacksonville,FL	S	Large City	15	53
33	67	Richmond,VA	S	Large City	111	69
34	68	Houston,TX	S	Large City	95	78
		EDC Model				
35		Philadelphia,PA	NE	Large City	112	-
36		Paterson,NJ	NE	Medium City	42	-
		Self-Sponsored Model				
37		Detroit,MI	NC	Large City	43	-
38	69	New York,NY	NE	Large City	20	13
39		Philadelphia,PA	NE	Large City	86	-
40		Portland,OR	W	Large City	45	-
41		San Diego,CA	W	Large City	71	-
		ILM Model				
42		New York,NY	NE	Large City	53	-
		SEDL Model				
43	70	Philadelphia,PA	NE	Large City	86	36
44		Tulare,CA	W	Small City	173	-
		Home-School Partnership Model				
45		New York,NY	NE	Large City	26	-
		California Process Model				
46		Los Angeles,CA	W	Large City	98	-
47		Ravenswood,CA	W	Small City	74	-
48		Lamont,CA	W	Rural Area	27	-
49		San Jose,CA	W	Large City	42	-
				Total Student Pop.	3210	1202

* Pooled Citywide NFT Population Figure for New York City

(Footnotes continued on next page)

Table A-2
(continued)

[1]NE = North Eastern United States

S = Southern United States

NC = North Central United States

W = Western United States

[2]Large City = 200,000 or more

Medium City = 50,000 to 199,999

Small City = 10,000 to 49,999

Rural Area = Less than 10,000

[3]All Data Envelopment Analysis study information refers to the Cohort II-K student population. II-K indicates that this group of students began their Program Follow Through experience in kindergarten. (This was also the only one of three Cohorts which had completed all of the grades from kindergarten through third grade at the time of our study for which site level information was available.) However, due to incomplete statistics along some DEA variable dimensions, some of the Cohort II-K PFT sites were not included in the DEA study. Specifically, Bank Street Model: Rochester,NY site; EDC Model: Chicago,IL site; and SEDL Model: St. Martin Parish,LA site were excluded from the DEA study student population. The actual Cohort II-K PFT population was 3,367 of which, as noted above, a set of 3,210 students were used in the DEA study. This exclusion of sites also extended to the NFT groups which were similarly reduced to 1,202 students.

[4]Two sets of NFT students groups were created in the original Program Follow Through study. One group was a local student set, usually in the same school system as the subject PFT site. The second group, and the one selected for the DEA study, was a "best matched" group, which may or may not have been located in the same school system or even the same geographical region. The NFT group which most nearly matched the PFT students of a given site along a number of demographic and initial performance dimensions was considered the "best match" for the latter. For several PFT sites the <u>same</u> "best matched" NFT group was used. The much smaller NFT student population total of 1,202 as compared to the PFT student total of 3,210 resulted. See also preceding footnote.

Table A-3

Unadjusted PFT Output Observations

Site #	Total Reading Scores, PHS* Y_1	Total Math Scores, PHS* Y_2	Total Coopersmith Scores, PHS* Y_3
1	54.53	58.98	38.16
2	24.69	33.89	26.02
3	36.41	40.62	28.51
4	14.94	17.58	16.19
5	7.81	6.94	5.37
6	12.59	16.85	12.84
7	17.06	16.99	17.82
8	20.29	30.64	33.16
9	26.13	29.80	26.29
10	46.42	51.59	35.20
11	39.80	37.73	30.29
12	37.84	47.85	25.35
13	26.48	31.36	26.54
14	10.31	10.86	7.47
15	14.39	18.30	14.33
16	32.94	36.03	38.19
17	17.25	20.80	12.07
18	27.55	38.19	20.44
19	41.12	43.80	36.54
20	29.43	42.63	23.34
21	37.46	51.02	27.44
22	19.40	25.18	16.52
23	39.88	47.72	38.97
24	25.72	30.81	16.54
25	24.88	25.27	22.43
26	31.62	40.78	31.16
27	31.31	38.32	25.03
28	21.00	21.30	18.30
29	6.51	7.02	6.16
30	11.64	15.26	15.68
31	12.58	15.90	14.42
32	4.59	6.16	4.99
33	43.76	46.64	39.10
34	32.38	38.55	31.05
35	34.64	45.46	39.22
36	11.52	15.14	13.91
37	15.96	19.21	15.30
38	9.91	12.30	7.22
39	30.44	33.53	29.80
40	22.63	25.24	17.15
41	24.41	27.16	25.30
42	23.11	22.67	17.56
43	21.82	31.45	27.54
44	63.92	79.67	63.11
45	9.47	11.92	8.85
46	33.94	39.18	34.61
47	29.42	35.10	28.42
48	7.70	11.02	9.02
49	12.17	16.03	15.82

* PHS = Per Hundred Students

Table A-4

Unadjusted NFT Output Observations

Site #	Total Reading Scores,PHS* Y_1	Total Math Scores,PHS* Y_2	Total Coopersmith Scores,PHS* Y_3
50	39.07	42.71	27.67
51	9.96	14.34	9.33
52	45.37	51.38	31.61
53	18.23	22.05	17.56
54	59.63	64.41	35.89
55	24.20	28.21	18.74
56	13.53	17.09	15.61
57	28.39	27.65	20.79
58	21.67	26.22	13.66
59	120.17	144.67	88.59
60	15.15	18.04	13.58
61	6.92	7.10	6.35
62	9.35	9.85	7.70
63	13.03	13.40	10.29
64	18.63	24.48	23.13
65	12.28	13.01	9.89
66	16.81	19.72	18.70
67	26.36	28.22	24.46
68	22.85	26.21	28.14
69	8.17	8.70	5.12
70	13.69	14.19	12.99

*PHS = Per Hundred Students

Table A-5

Unadjusted PFT Input Observations

Site #	Education Level of Mother, PHS* X_1	Occupation Index, PHS* X_2	Parental Visit Index, PHS* X_3	Counseling Index, PHS X_4	Number of Teachers X_5
1	86.13	16.24	48.21	49.69	9
2	29.26	10.24	41.96	40.65	5
3	43.12	11.31	38.19	35.03	9
4	24.96	6.14	24.81	25.15	7
5	11.62	2.21	6.85	6.37	4
6	11.88	4.97	18.73	18.04	4
7	32.64	6.88	28.10	25.45	7
8	20.79	12.97	54.85	52.07	8
9	34.40	11.04	38.16	42.40	8
10	61.74	14.50	49.09	42.92	9
11	52.92	11.67	39.48	39.64	5
12	36.00	10.15	37.80	39.52	5
13	39.20	10.80	41.04	41.12	7
14	14.6	2.88	9.64	11.14	3
15	4.29	5.42	21.45	17.27	5
16	27.25	14.17	56.46	55.26	9
17.	22.63	4.43	15.40	15.00	2
18	28.00	7.61	28.73	27.04	9
19	53.56	13.70	53.04	49.85	7
20	25.42	9.05	29.69	31.74	4
21	31.57	10.08	39.34	40.57	6
22	16.34	5.84	20.89	22.10	4
23	44.28	14.14	56.70	52.27	11
24	19.74	6.43	24.20	25.66	3
25	24.40	8.05	33.42	31.29	7
26	41.40	11.70	44.01	46.35	7
27	27.20	9.38	37.80	31.55	4
28	23.92	7.12	25.58	29.01	3
29	10.62	2.55	10.10	9.09	4
30	12.48	6.14	23.13	22.46	6
31	19.32	5.89	24.01	24.74	6
32	6.30	1.93	7.11	7.68	4
33	46.62	14.65	65.71	57.49	10
34	38.95	12.82	47.02	48.92	9
35	61.60	15.56	53.98	50.29	6
36	31.08	6.26	22.18	21.96	4
37	19.35	6.68	22.61	23.31	4
38	11.20	3.08	9.90	10.06	2
39	34.40	11.61	41.79	41.79	5
40	35.55	6.48	21.69	21.69	6
41	30.53	9.30	35.50	35.14	8
42	25.44	7.10	26.81	26.23	3
43	26.66	11.43	41.36	44.63	6
44	39.79	22.49	84.77	76.12	11
45	8.32	3.64	12.92	13.13	2
46	59.78	13.52	48.80	49.69	15
47	39.22	10.06	37.00	38.33	4
48	3.28	3.18	13.12	12.71	5
49	7.14	5.29	23.10	19.06	8

*PHS = Per Hundred Students

Table A-6

Unadjusted NFT Input Observations

Site #	Education Level of Mother, PHS* X_1	Occupation Index, PHS* X_2	Parental Visit Index, PHS* X_3	Counseling Index, PHS* X_4	Number of Teachers X_5
50	68.16	12.28	33.58	34.64	15
51	11.88	3.59	13.41	13.82	8
52	55.30	11.53	36.73	35.78	6
53	16.20	7.02	26.94	26.30	9
54	82.45	15.52	45.00	44.23	13
55	15.81	6.93	23.91	23.61	7
56	4.65	5.50	20.91	23.39	5
57	41.25	8.41	26.23	25.24	10
58	10.44	5.22	17.10	18.93	3
59	139.65	35.03	119.56	130.83	22
60	16.28	4.81	18.20	18.98	5
61	12.06	2.59	8.74	8.17	5
62	4.20	2.64	9.89	11.25	2
63	19.44	3.83	12.87	13.23	5
64	28.38	8.91	30.95	33.33	8
65	13.50	3.61	15.60	12.39	4
66	23.32	7.10	24.96	28.56	22
67	27.60	9.38	32.29	34.01	20
68	11.70	10.53	37.67	43.60	8
69	4.68	1.85	6.22	5.46	5
70	10.44	4.82	17.13	18.21	9

*PHS = Per Hundred Students

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER MSRR 432	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A DATA ENVELOPMENT ANALYSIS APPROACH TO EVALUATION OF THE PROGRAM FOLLOW THROUGH EXPERIMENT IN U.S. PUBLIC SCHOOL EDUCATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report November 1978
		6. PERFORMING ORG. REPORT NUMBER MSRR 432
7. AUTHOR(s) A. Charnes W. W. Cooper E. Rhodes		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0616 N00014-76-C-0932 SOC76-15876
9. PERFORMING ORGANIZATION NAME AND ADDRESS Graduate School of Industrial Administration Carnegie-Mellon University Pittsburgh, Pennsylvania 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 047-021
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 434) Arlington, Virginia 22217		12. REPORT DATE November 1978
		13. NUMBER OF PAGES 59
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Efficiency, Program Efficiency, Managerial Efficiency, Decision Making Units, ProgramFollow Through, Educational Outputs, Mathematical Programming, Linear Programming, Duality Relations, Extremal Relations, Efficiency Frontiers, Isoquants, Production Possibility Surfaces, Information Measures, Regression, Simultaneous Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method called Data Envelopment Analysis (DEA) is used to decompose the efficiency of Decision Making Units (DMU's) into two parts: (1) a component resulting from managerial decisions and (2) a component resulting from constraints (called programs) under which management operates. The DEA approach accomplishes this by enveloping the input-out-ut observations with extremal relations developed in terms of a specified nonlinear programming model (and/or its linear programming equivalent). Differences between the observations and the program specific envelopes ~~--- called α-envelopes --- are~~ (continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

imputed to managerial inefficiencies. An inter-program envelope is then constructed from 2 or more such α-envelopes and used to identify "program" inefficiencies, which are the inefficiencies that remain after the previously determined managerial inefficiencies have been eliminated. Numerical illustrations accompanied by suggested tests of a probabilistic/information theoretic character are provided by means of recently released data from "Program Follow Through." Designed as a study of possible ways of reenforcing or extending Program Head Start - an ongoing pre-school program for disadvantaged children -- the Program Follow Through experiment provides data on agreed upon inputs and outputs for both PFT (Program Follow Through) and matched NFT (Not Follow Through) participants in various parts of the U.S. Only a subset of the variables from the Follow Through experiment are used. Hence the numerical example utilized here is best regarded as only illustrative. Although the results are adverse to PFT, the DEA approach also opens new ways of profiting from the results of such experiments by examining combinations of the underlying components. These kinds of possibilities are also described in this paper.